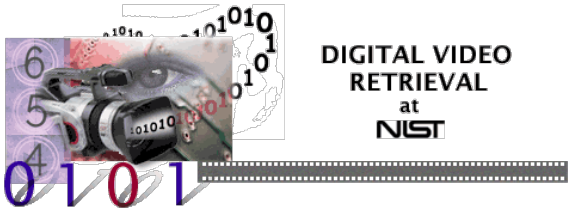
2014 TRECVID Workshop: **Surveillance Event Detection** (SED) Retrospective + Interactive (rSED+iSED) Task Overview

National Institute of Standards and Technology (NIST)

Martial Michel, PhD

David Joy

November 12, 2014



Home Office Scientific Development Branch | CPNI Centre for the Protection of National Infrastructure

NIST National Institute of Standards and Technology

About the SED Evaluation

- Surveillance Event Detection Motivation

  SED addresses the need for the advancement of technologies that can perform automatic detection of events in large amounts of surveillance quality video

- Identify each detected event observation by:
  - The ***temporal extent*** *(beginning and end frames)*
  - A ***decision score***: a numeric score indicating how likely the event observation exists with more positive values indicating more likely observations (normalized)
  - An ***actual decision***: a Boolean value indicating whether or not the event observation should be counted for the primary metric computation

SED Tasks

Retrospective SED (rSED): Given a textual description of an *observable event of interest,* **automatically detect** all occurrences of the event in a non-segmented corpus of video

- Requires application of several Computer Vision techniques
- Involves subtleties that are readily understood by humans, difficult to encode for machine learning approaches
- Can be complicated due to clutter in the environment, lighting, camera placement, traffic, etc.

Interactive SED (iSED): Given a textual description of an *observable event of interest,* at **test time allow a searcher 25 minutes to filter incorrect event detections** from the rSED task

- in a non-segmented corpus of video
- SED remains a difficult task for humans and systems
- Interactive/relevance feedback have been effectively employed in other related tasks

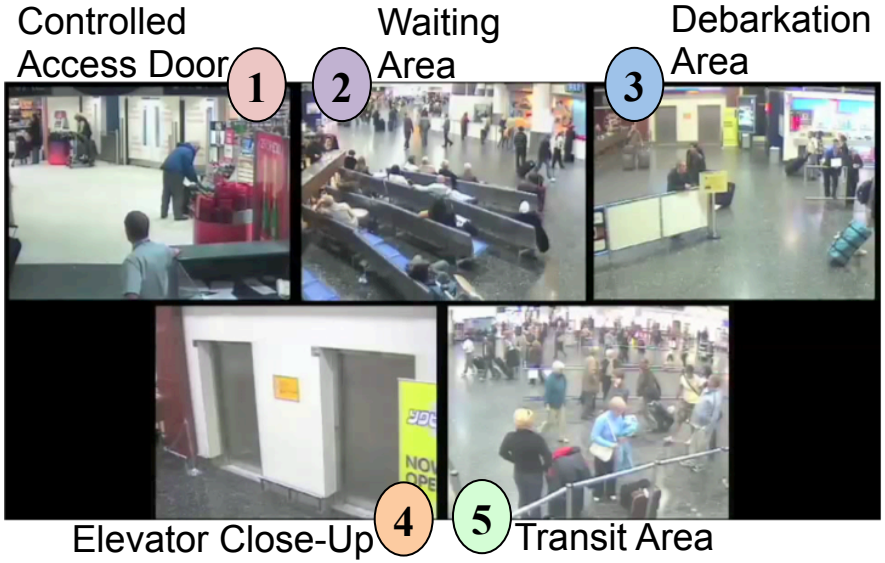
Events of Interest

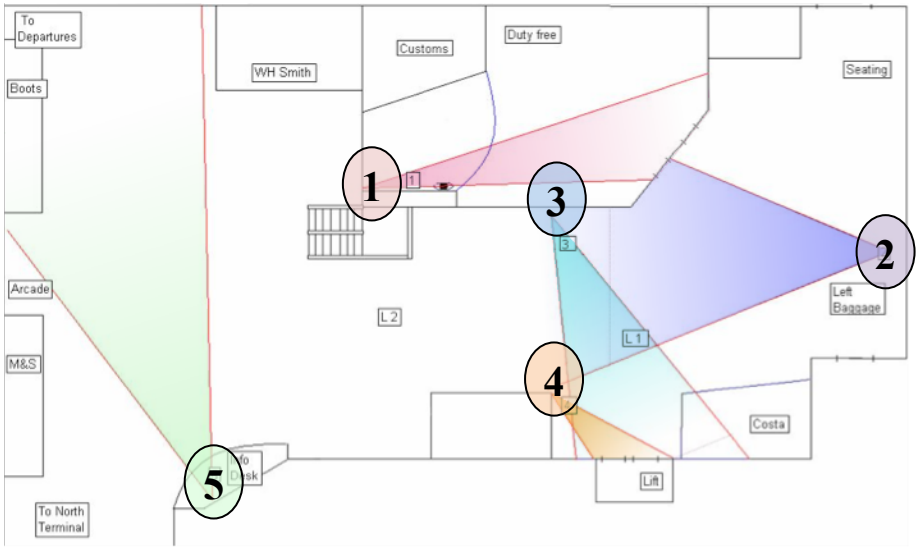
Single Person events	
PersonRuns	Someone runs
Pointing	Someone points
Single Person + Object events	
CellToEar	Someone puts a cell phone to his/her head or ear
ObjectPut	Someone drops or puts down an object
Multiple People events	
Embrace	Someone puts one or both arms at least part way around another person
PeopleMeet	One or more people walk up to one or more other people, stop, and some communication occurs
PeopleSplitUp	From two or more people, standing, sitting, or moving together, communicating, one or more people separate themselves and leave the frame

Evaluation Source Data

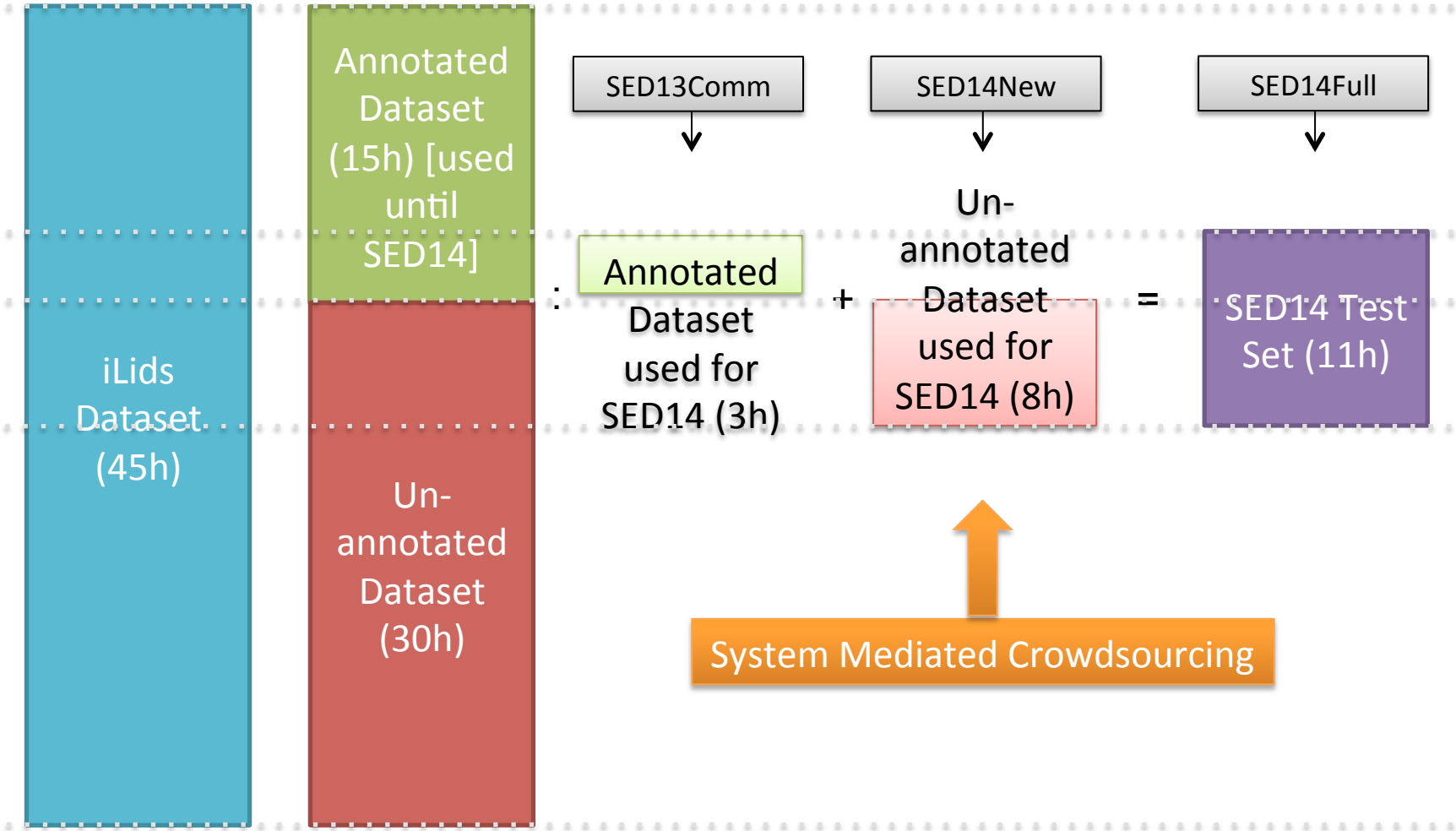
Imagery Library for Intelligent Detection Systems

- UK Home Office collected CCTV video from 5 camera views at a busy airport
- Development Set
  - 100 hours of video
  - 10 events annotated on 100% of the data
- Evaluation Set (SED '09 '10 '11 '12 '13)
  - "iLIDS Multiple Camera Tracking Scenario Training set"
  - An identified 15-hours of the 45-hour set evaluated
  - 10 events annotated on 1/3 of the data
    - 7 events evaluated
- Evaluation Set (SED '14)
  - Subset of 11hours of the 45-hour iLIDS data set evaluated (3h common to SED13 + 8h new data)





SED14 Evaluation Set



SED14New System Mediated Crowdsourcing: Bootstrap level 1 (of 2)

1. Obtain system output from top past performers on SED14New (**BUPT, CMU, IBM, PKU**)
2. Calculate agreement for each event instance
3. Review
4. Generate “bootstrap level 1”

↓Event | Agree→	25%
CellToEar	307
Embrace	2837
ObjectPut	1233
PeopleMeet	1906
PeopleSplitUp	489
PersonRuns	948
Pointing	18250
Grand Total	**25970**

"Event Instance Review and Annotation" software

Browse... MCTTR0705c_adjudicated(2).csv Save (Right click to save as)

Event Description

PeopleMeet

One or more people walk up to one or more other people, stop, and some communication occurs.

Start: The, first communication between any member of one group to a member of the other group.

End: The earliest time when the two groups are nearest to each other after the communication has occurred.

S: 5161 <- **5161** -> E: 5221 PeopleMeet

<< -1 +1 >> [P]lay [P]ause [N]o [R]eset [Y]es << -1 +1 >>

EventType	ID	Start	End	Agree
PeopleMeet	29	2971	3031	25%
PeopleMeet	30	3091	3151	25%
PeopleMeet	31	3481	3541	25%
PeopleMeet	32	3541	3601	25%
PeopleMeet	33	3601	3661	25%
PeopleMeet	34	3661	3721	25%
PeopleMeet	36	4741	4801	25%
PeopleMeet	37	4801	4861	25%
PeopleMeet	38	5161	5221	25%
PeopleMeet	39	5941	6001	25%
PeopleMeet	40	6151	6211	25%
PeopleMeet	41	6241	6301	25%
PeopleMeet	42	6751	6811	25%
PeopleMeet	43	6841	6901	25%
PeopleMeet	44	691	751	25%
PeopleMeet	45	6931	6991	25%
PeopleMeet	46	6991	7051	25%
PeopleMeet	47	7051	7111	25%
PeopleMeet	48	7201	7261	25%
PeopleMeet	49	7321	7381	25%
PeopleMeet	50	7381	7441	25%

Default Sort
Add Instance
Delete Selected Instance

Some buttons have [hotkeys], up and down arrows navigate annotations.
Left click and hold to draw a bounding box for the start frame, press [c] to clear the bounding box.
[s] and [e] set the start and end frame (respectively) to be equal to the current frame.
Left and Right arrow keys move the current frame irrespective of the instance bounds. The number of frames moved changes with certain modifier keys (default: 1, +shift: 10, +alt: 50).

SED14New System Mediated Crowdsourcing: Bootstrap level 2 (of 2)

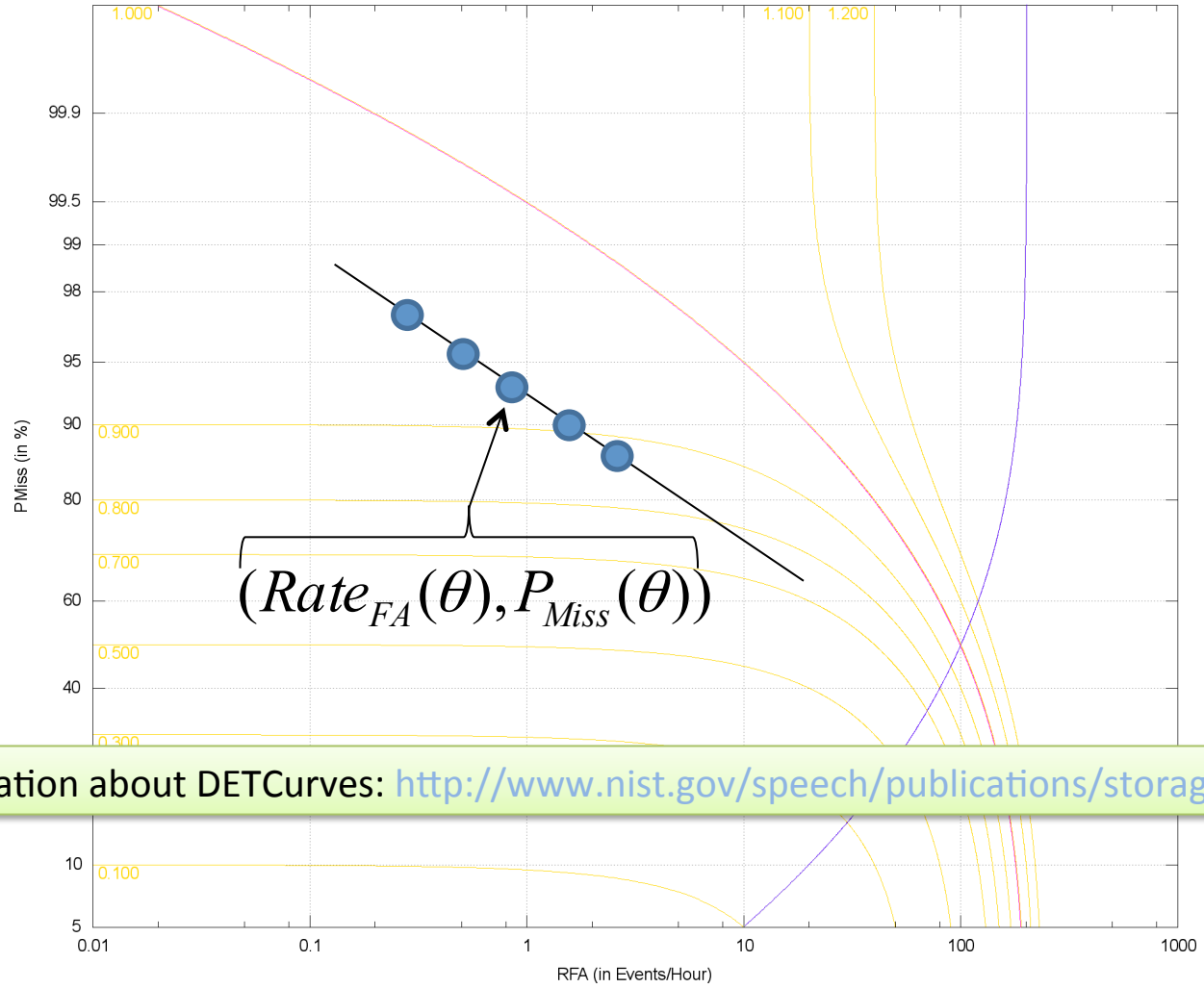
1. Use SED14 top systems (1 per site/per event, independent of task)
2. Remove bootstrap level 1 event instances
3. Calculate agreement for each event instance
4. Review (Include extra events found during review process)
5. Generate "bootstrap level 2"

   (Add new events to bootstrap level 1 reference)

↓Event | Agree→	25%	50%	75%	100%	Reviewed
CellToEar	86				86
Embrace	284	18	2		304
ObjectPut	139	4			143
PeopleMeet	423	329	42	2	796
PeopleSplitUp	647	34			681
PersonRuns	269	13	1		283
Pointing	276	14			290
Grand Total	**2124**	**412**	**45**	**2**	**2583**

SED Error Visualization

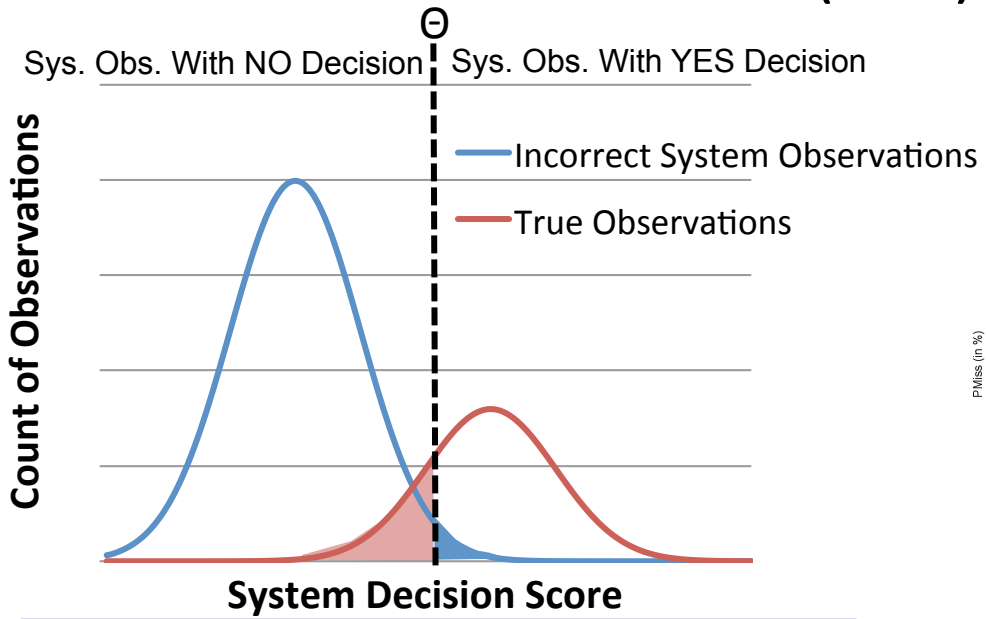
Detection Error Tradeoff (DET) Curves ($Prob_{Miss}$ vs. $Rate_{FA}$)



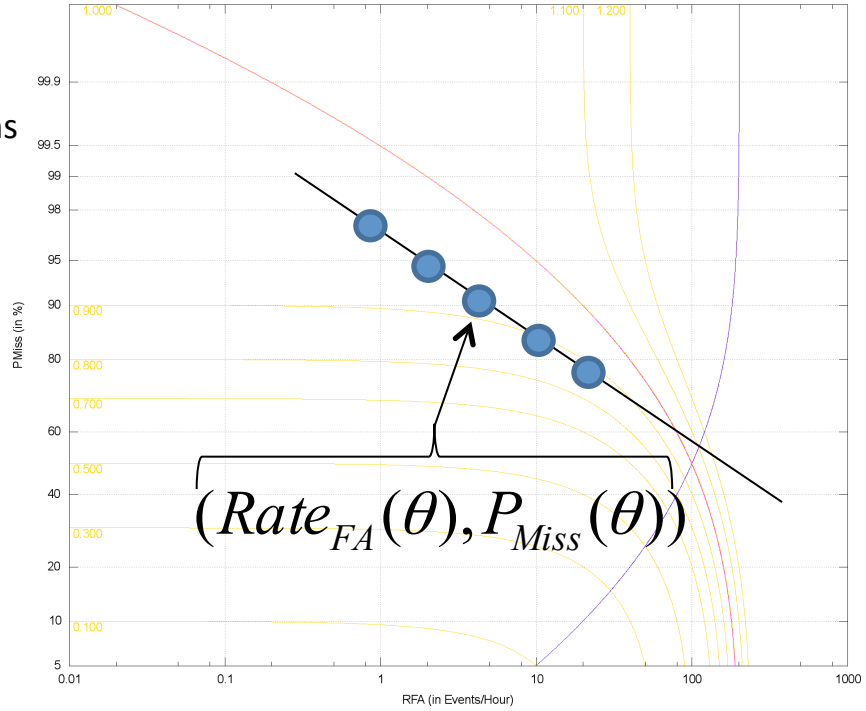
For more information about DETCurves: http://www.nist.gov/speech/publications/storage_paper/det.pdf

SED Error Visualization

Detection Error Tradeoff (DET) Curves (*Prob*$_{Miss}$ vs. *Rate*$_{FA}$)



Compute Rate$_{FA}$ and P$_{Miss}$ for all Θ



$$MinNDCR(\theta) = \underset{\theta}{\operatorname{argmin}}\left[P_{Miss}(\theta) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(\theta)\right]$$

$$ActNDCR(Act.Dec.) = P_{Miss}(Act.Dec.) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(Act.Dec.)$$

For more information about DETCurves: http://www.nist.gov/speech/publications/storage_paper/det.pdf

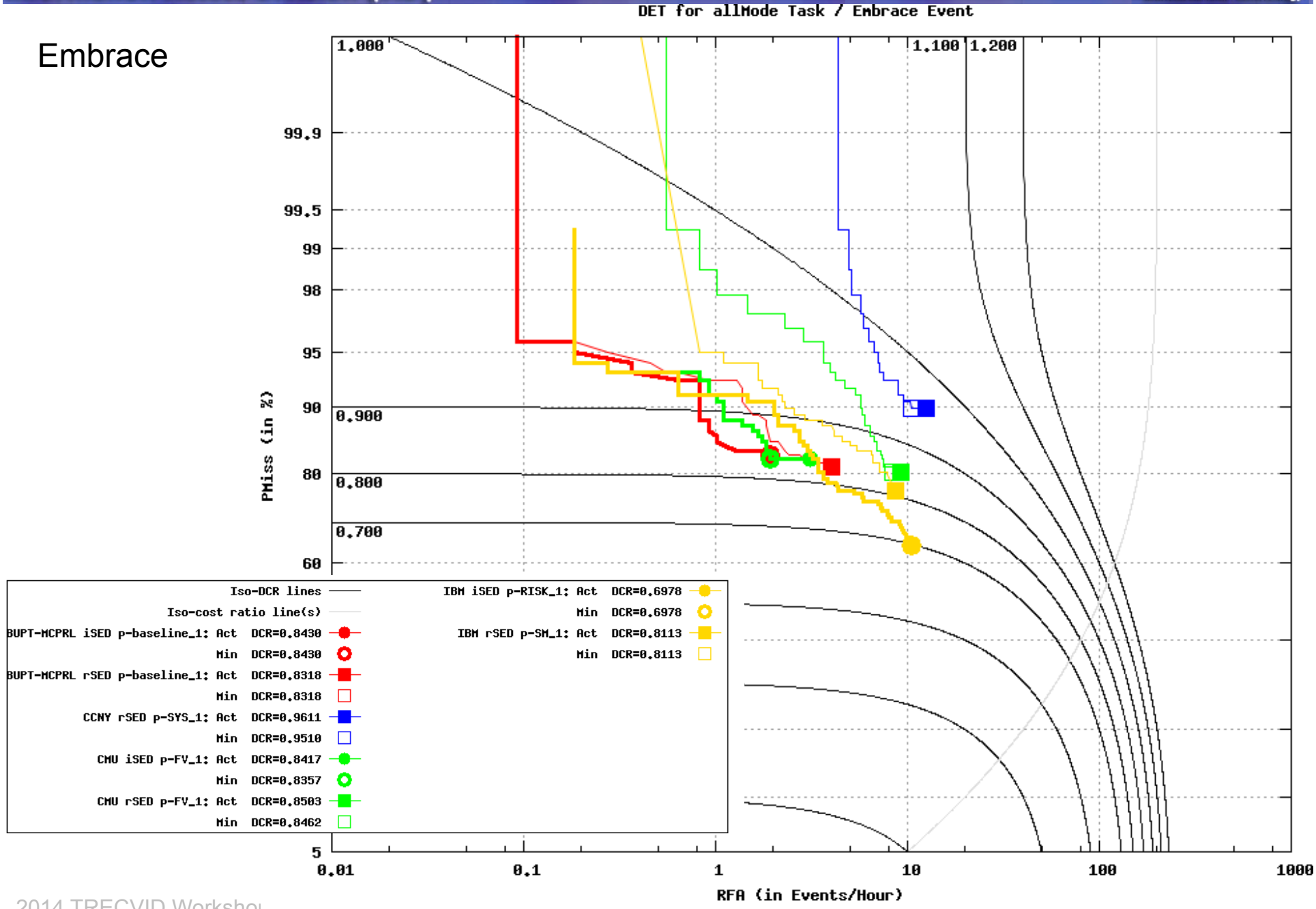
4 SED 2014 Participants

(with number of systems per event)

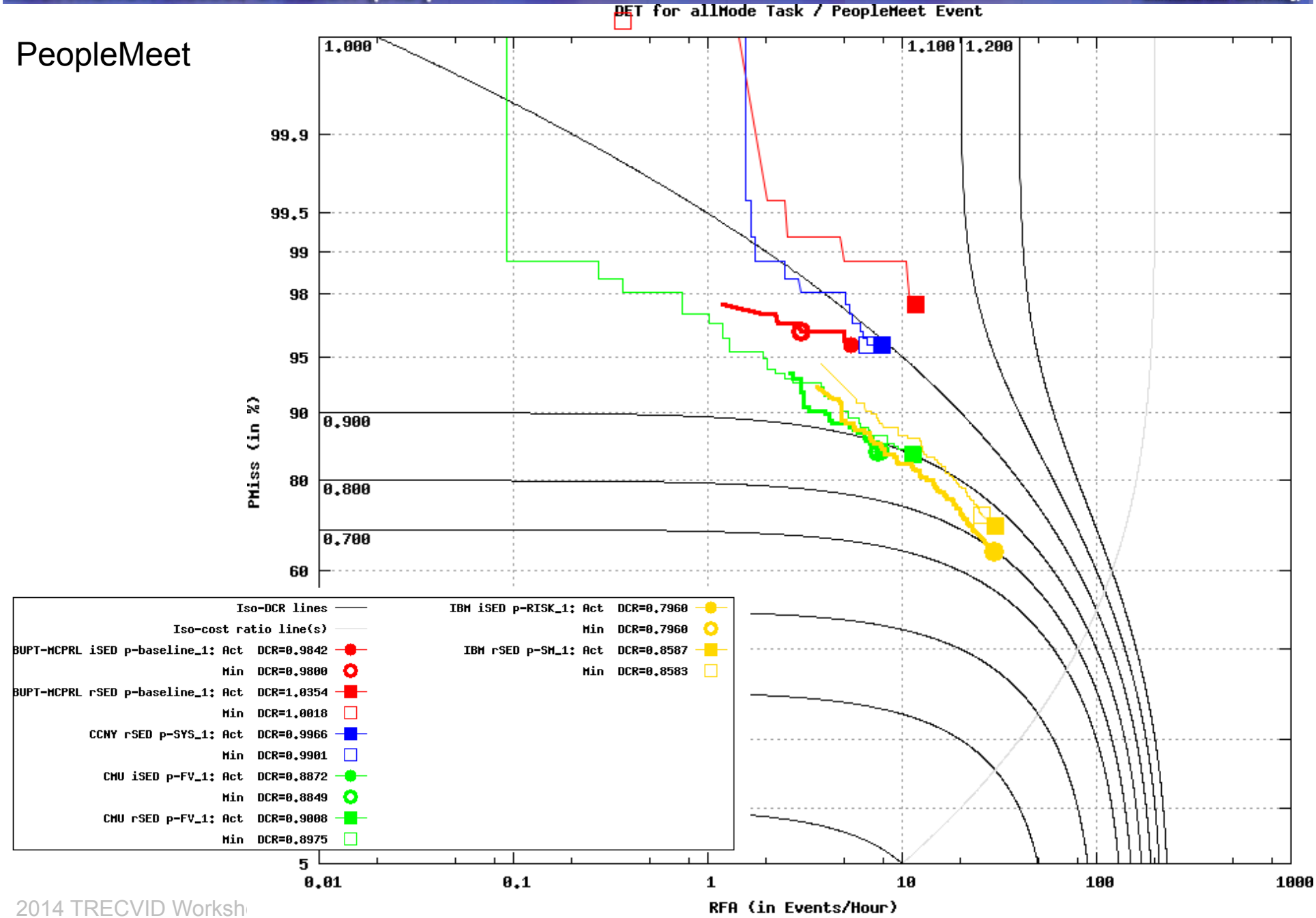
		Single Person				Person + object				Multiple People					
		PersonRuns		Pointing		CellToEar		ObjectPut		Embrace		PeopleMeet		PeopleSplitUp	
		iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED
7 years in a row	Carnegie Mellon University [CMU]	3	3	3	3	3	3	3	3	3	3	3	3	3	3
6 years in a row	Multimedia Communication and Pattern Recognition Labs, Beijing University of Posts and Telecommunications [BUPT-MCPRL]	2	2	2	2					2	2	2	2	2	2
3 years in a row	IBM Thomas J. Watson Research Center [IBM]	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	The City College of New York Media Lab [CCNY]		2		2		2		2		2		2		2
		6	8	6	8	4	6	4	6	6	8	6	8	6	8

Total iSED Runs	38
Total rSED Runs	52

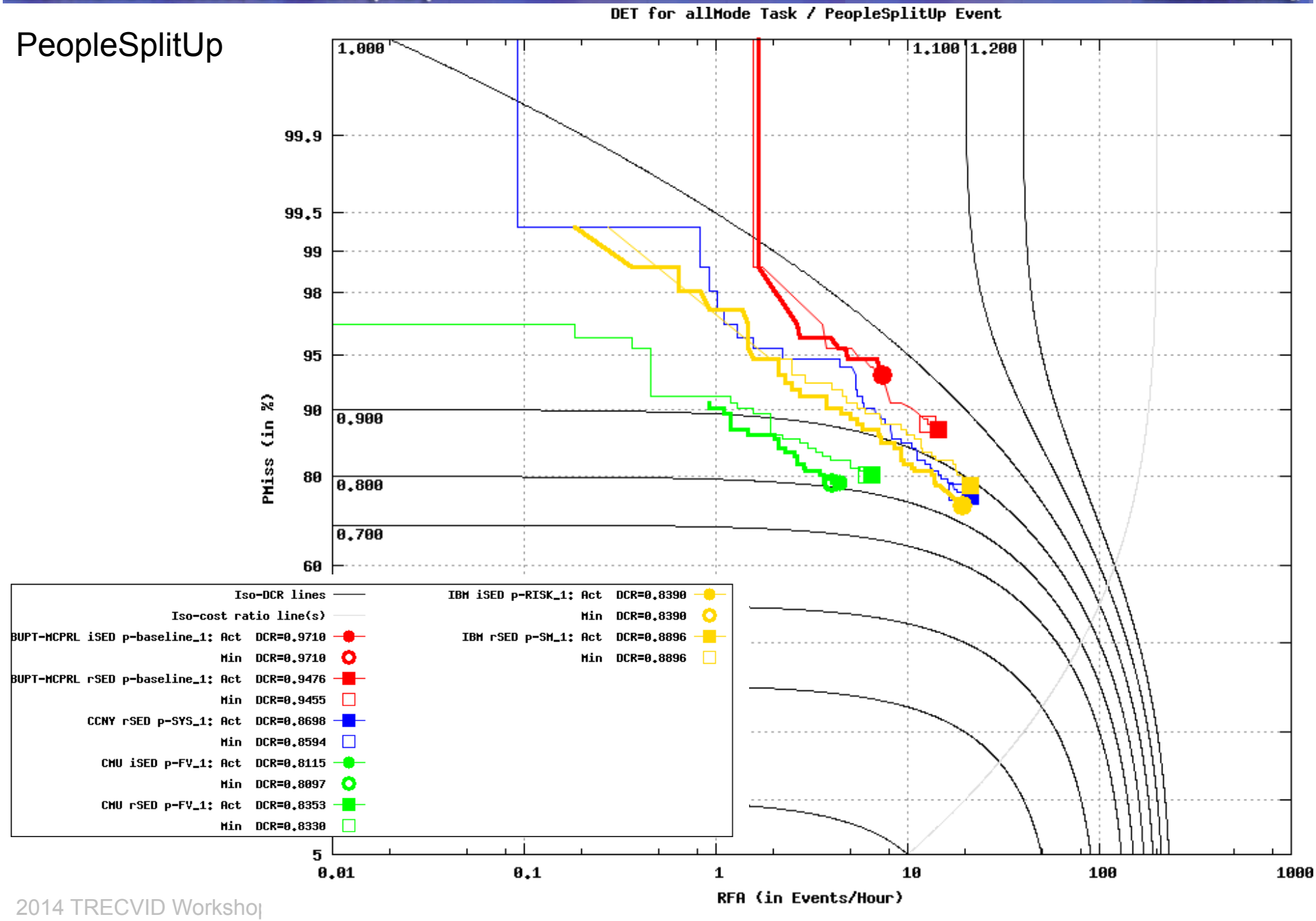
Embrace


DET for allMode Task / Embrace Event
1.000
1.100
1.200
99.9
99.5
99
98
95
90
0.900
80
0.800
0.700
60
PMiss (in %)
5
0.01
0.1
1
10
100
1000
RFA (in Events/Hour)
Iso-DCR lines
Iso-cost ratio line(s)
BUPT-MCPRL iSED p-baseline_1: Act DCR=0.8430
Min DCR=0.8430
BUPT-MCPRL rSED p-baseline_1: Act DCR=0.8318
Min DCR=0.8318
CCNY rSED p-SYS_1: Act DCR=0.9611
Min DCR=0.9510
CMU iSED p-FV_1: Act DCR=0.8417
Min DCR=0.8357
CMU rSED p-FV_1: Act DCR=0.8503
Min DCR=0.8462
IBM iSED p-RISK_1: Act DCR=0.6978
Min DCR=0.6978
IBM rSED p-SM_1: Act DCR=0.8113
Min DCR=0.8113

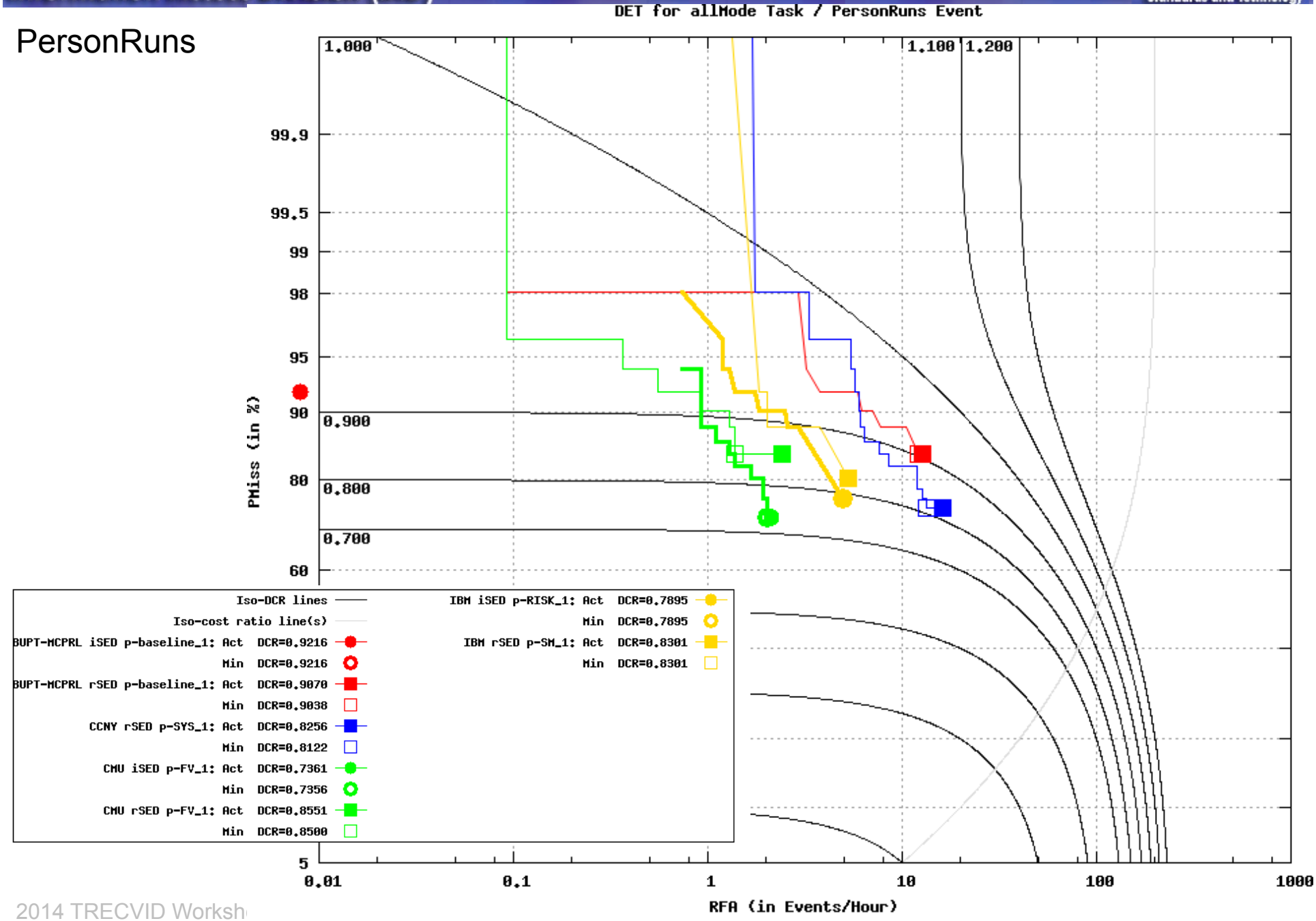
PeopleMeet


DET for allMode Task / PeopleMeet Event
1.000
1.100
1.200
99.9
99.5
99
98
95
90
80
60
5
0.900
0.800
0.700
PMiss (in %)
0.01
0.1
1
10
100
1000
RFA (in Events/Hour)
Iso-DCR lines
Iso-cost ratio line(s)
BUPT-MCPRL iSED p-baseline_1: Act DCR=0.9842
Min DCR=0.9800
BUPT-MCPRL rSED p-baseline_1: Act DCR=1.0354
Min DCR=1.0018
CCNY rSED p-SYS_1: Act DCR=0.9966
Min DCR=0.9901
CMU iSED p-FV_1: Act DCR=0.8872
Min DCR=0.8849
CMU rSED p-FV_1: Act DCR=0.9008
Min DCR=0.8975
IBM iSED p-RISK_1: Act DCR=0.7960
Min DCR=0.7960
IBM rSED p-SM_1: Act DCR=0.8587
Min DCR=0.8583

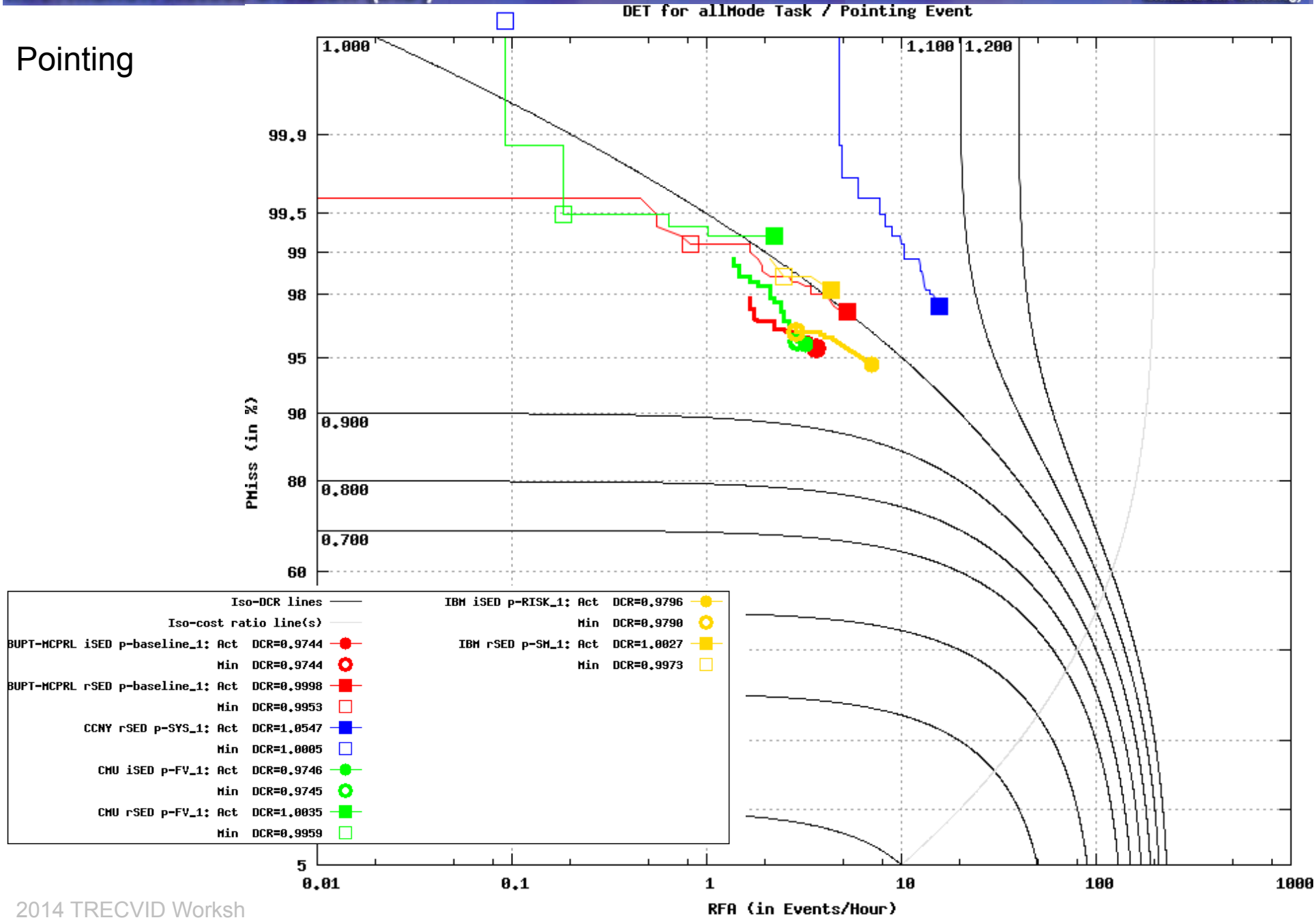
PeopleSplitUp



PersonRuns


DET for allMode Task / PersonRuns Event
PMiss (in %)
RFA (in Events/Hour)
1.000
1.100
1.200
0.900
0.800
0.700
99.9
99.5
99
98
95
90
80
60
5
0.01
0.1
1
10
100
1000
Iso-DCR lines
Iso-cost ratio line(s)
BUPT-MCPRL iSED p-baseline_1: Act DCR=0.9216
Min DCR=0.9216
BUPT-MCPRL rSED p-baseline_1: Act DCR=0.9070
Min DCR=0.9038
CCNY rSED p-SYS_1: Act DCR=0.8256
Min DCR=0.8122
CMU iSED p-FV_1: Act DCR=0.7361
Min DCR=0.7356
CMU rSED p-FV_1: Act DCR=0.8551
Min DCR=0.8500
IBM iSED p-RISK_1: Act DCR=0.7895
Min DCR=0.7895
IBM rSED p-SM_1: Act DCR=0.8301
Min DCR=0.8301

Pointing


DET for allMode Task / Pointing Event
PMiss (in %)
RFA (in Events/Hour)
1.000
1.100
1.200
0.900
0.800
0.700
99.9
99.5
99
98
95
90
80
60
40
20
10
5
0.01
0.1
1
10
100
1000
Iso-DCR lines
Iso-cost ratio line(s)
BUPT-MCPRL iSED p-baseline_1: Act DCR=0.9744
Min DCR=0.9744
BUPT-MCPRL rSED p-baseline_1: Act DCR=0.9998
Min DCR=0.9953
CCNY rSED p-SYS_1: Act DCR=1.0547
Min DCR=1.0005
CMU iSED p-FV_1: Act DCR=0.9746
Min DCR=0.9745
CMU rSED p-FV_1: Act DCR=1.0035
Min DCR=0.9959
IBM iSED p-RISK_1: Act DCR=0.9796
Min DCR=0.9790
IBM rSED p-SM_1: Act DCR=1.0027
Min DCR=0.9973

Conclusion

- System Mediated REF generation a possibility
  - Participation requiring annotation of a selection of the reference video data on a set of events
  - Reference extended by post adjudication
  - Requires a lot of human time for review process (despite pre-search by systems)
  - Some events are not found by computers
    - Unless a human does an extra pass on video, those events that are not detected by any systems will be missed

Future of SED Evaluation

- SED15 to reuse same test data set as SED14
  - Discussion: New Events ?

✓Work on better reference

✓Bounding boxes to review event occurrence invaluable for crowded scene